

SVEUČILIŠTE U ZAGREBU  
FILOZOFSKI FAKULTET  
ODSJEK ZA INFORMACIJSKE ZNANOSTI  
ODSJEK ZA ZAPADNOSLAVENSKE JEZIKE I KNJIŽEVNOSTI  
(BOHEMISTIKA)

Matija Latin

**Valencijski rječnici u češkoj i hrvatskoj leksikografiji**

Diplomski rad

Mentorice:

doc. dr. sc. Nives Mikelić Preradović

prof. dr. sc. Dubravka Sesar

Zagreb, veljača 2013.

## Sadržaj:

I.	Uvod.....	3
II.	Termin „valencija“ i teorija valentnosti.....	4
III.	Teoretski temelji VALLEX-a i CROVALLEX-a .....	5
IV.	Primjena valencijskih rječnika: .....	7
V.	Valencijski rječnici češkog jezika .....	11
V.1.	VALLEX 1.0 .....	11
V.2.	VALLEX 1.0. – logička struktura podataka.....	12
V.3.	VALLEX 2.0 i VALLEX 2.5 .....	23
VI.	CROVALLEX - valencijski rječnik hrvatskog jezika .....	26
VII.	Usporedbena analiza valentnosti glagola u češkom i hrvatskom jeziku .....	26
VIII.	Pravila realizacije dopuna .....	34
IX.	Zaključak .....	36
X.	Literatura.....	37
XI.	Sažetak na hrvatskom jeziku .....	38
XII.	Sažetak na češkom jeziku .....	39

## I. Uvod

Valencijski rječnici možda i nisu toliko poznati široj javnosti, no vrlo su korisni i zanimljivi svakome tko se bavi obradom prirodnog jezika. Od velike koristi mogu biti i neizvornim govornicima, koji tek uče strani jezik. Upravo u toj situaciji sam se i ja prvi puta susreo s valencijskim rječnikom glagola u češkom jeziku, koji sam redovito koristio tijekom cijelog studija češkog jezika i na kojem se ovaj diplomski rad u konačnici i temelji.

U ovom radu će biti opisana kratka povijest nastanka valencijskih rječnika te same teorije valentnosti na kojoj se oni temelje. Nadalje, bit će navedeni i opisani najznačajniji valencijski rječnici glagola u češkom jeziku te valencijski rječnik glagola u hrvatskom jeziku, koji je nastao na istim temeljima.

Završna poglavlja ovog diplomskog rada bit će posvećena istraživanju sličnosti i razlika među valencijskim čeških i hrvatskih glagola. Kao cilj rada postavljeno je utvrđivanje pravila po kojima se određeni valencijski okviri u češkom jeziku mogu direktno prevoditi na hrvatski jezik.

## II. Termin „valencija“ i teorija valentnosti

Teorija valentnosti ogranak je gramatike zavisnosti, koja je nastala u Francuskoj sredinom 20. stoljeća. Što je zapravo valentnost? Sam termin valencije je u sklopu kemije 1852. godine definirao Edward Frankland, koji ga je tada opisao kao svojstvo atoma da na sebe veže određen broj drugih atoma. Nešto više od sto godina kasnije, 1959. godine, Lucien Tesnière isti termin uvodi u lingvistiku i koristi ga u kontekstu sintaktičke analize rečenice (Žabokrtský 2005, str. 1).

Isto kao što atom može uz sebe vezati određeni broj drugih atoma, tako i glagol, prema Tesnieren, može uz sebe vezati podređene elemente, koje on naziva *aktantima*. U lingvistici se tako aktantom naziva biće ili stvar koja bilo na koji način sudjeluje u radnji, koji se obvezno mora pojaviti i koji predstavlja cjelinu s glagolom.

Tesnière je vršitelja, trpitelja i primatelja radnje nazvao prvim, drugim i trećim aktantom, od kojih se prvi odnosi na subjekt, drugi na direktni objekt, a treći se odnosi na entitet u čiju se korist događa radnja glagola. Tesnière je također definirao i drugi važni tip komplementa, a to su dodatci (koje on naziva *cirkumstantima*), odnosno elementi koji nisu nužno uvjetovani glagolom u rečenici i njihov broj nije unaprijed utvrđen, ali pokazuju mjesto, vrijeme, način i ostala obilježja radnje.

U lingvistici tako pojam valencije počinje označavati sposobnost nositelja valentnosti (to može biti glagol, imenica ili pridjev, no ovaj rad je usredotočen prvenstveno na valentnost glagola) da u rečenici otvara određen broj mjesta koja se morfološki i semantički moraju popuniti određenim komponentama (Mikelić-Preradović 2008: str 81.).

Nadalje, Tesnière glagole dijeli na nevalentne (neosobne, npr. grmjeti, kišiti), jednovalentne (neprijelazne, npr. bdijeti, spavati), dvovalentne (prijelazne, npr. postati, ubrzati) i trovalentne (prijelazne s dva objekta, npr. pitati, staviti). (Gulešić-Machata, Čilaš-Mikulić, Udier 2011: str. 24.).

Elementi koji dolaze uz glagol, dijele se na dopune (inner participants) i dodatke (free modifications), s tim da dopune mogu biti obvezne i izborne, a izborne dopune opet mogu biti tipične ili netipične. Izostavljanje obvezne dopune iz rečenice čini rečenicu negramatičnom (Mikelić-Preradović 2008: str. 81.).

### III. Teoretski temelji VALLEX-a i CROVALLEX-a

Teoretska osnova za izradu valencijskih rječnika VALLEX-a i CROVALLEX temelji se na funkcijskom generativnom opisu (FGD, *Functional Generative Description*). FDG je sustav opisa prirodnog jezika, koji je u Pragu razvila grupa matematičkih lingvista predvođenih Sgallom. FGD karakterizira stratifikacijski pristup temeljen na zavisnosti i dvorazinskom sintaktičkom označavanju:

- analitička razina – površinski sintaktički prikaz
- tektogramatička razina – pozadinska struktura

Analitička razina predstavlja površinski sintaktički prikaz, dok tektogramatička razina obuhvaća pozadinsku strukturu (Mikelić-Preradović 2008: str. 81.).

Glagoli se u FGD-u, kao i u drugim sličnim opisima, smatraju središtem rečenice te je zbog toga najveći naglasak stavljen na sintaktičke osobine glagola, tj. na formulaciju kriterija za utvrđivanje pojedinih komplemenata, koji tvore valencijski okvir te na utvrđivanje njihovih osobina, što je pak naročito bitno želi li se opis iskoristiti za automatsku obradu.

Glagolskim komplementima se smatraju izrazi koji mogu razvijati glagol pa tako valencijski okvir sačinjavaju svi komplementi koji razvijaju određeni glagol u određenom smislu. Sam valencijski okvir dijeli se na *inner participants* (dopune) i *free modifications* (dodatke) glagola.

FGD razlikuje pet tipova dopuna, za koje vrijedi pravilo da je određena kombinacija dopuna karakteristična za određen glagol te da se svaka dopuna može u ulozi komplementa za određeni glagol pojaviti samo jednom. Dopune tako mogu biti *Actor* (ACT), *Patient* (PAT), *Addressee* (ADDR), *Origin* (ORIG) i *Effect* (EFF).

Primjer:

*Matka*.ACT      *předělala*      *dětem*.ADDR.      *loutku*.PAT  
*z Kašpárka*.ORIG      *na čerta*.EFF

(Majka je djeci prepravila lutku iz Kašparka u ďavla.)

Postoje i glagoli koji u svom valencijskom okviru nemaju niti jednu dopunu, kao na primjer glagol *pršet*.

*Venku prší.*

(Vani kiši.)

Dodaci pak označavaju npr. mjesto, vrijeme, način, sredstvo, uvjet itd. i mogu se pojavljivati kod bilo kojeg glagola te isti glagol može imati više dodataka istog tipa:

*V Praze se sejdeme na Hlavním nádraží u pokladen.*

(U Pragu ćemo se naći na Glavnom kolodvoru kod blagajni.)

*Včera přišel večer domů pozdě.*

(Jučer navečer je došao kući kasno.)

*Kvůli dešti musel čekat pod střechou, protože neměl deštník.*

(Zbog kiše je morao čekati pod krovom jer nije imao kišobran.)

Dopune i dodaci mogu kod pojedinih glagola biti *obvezni* (tj. za određeni glagol obavezno prisutni u dubinskoj strukturi rečenice) ili *izborni*. (Loptaková, Žabokrtský, 2002: str. 2.).

#### IV. Primjena valencijskih rječnika:

Gulešić-Machata, Čilaš-Mikulić i Udier (2011: str. 24.) ističu da su podaci o tome kojim riječima pojedini glagol otvara mjesto, odnosno kojim i kakvim riječima može biti dopunjavani u kojem značenju, važni i izvornim govornicima i jezičnim stručnjacima, no još važniji su neizvornim govornicima koji uče određeni jezik jer je bez znanja o valentnosti pojedini glagol nemoguće upotrijebiti, odnosno proizvesti njime rečenicu određena značenja koje se želi izraziti. Autorice navode primjer triju rečenica s hrvatskim glagolom *ići* kod kojih i kod izvornih govornika još postoje dvojbe jesu li pravilne i prihvatljive:

- 1) Idem liječniku.
- 2) Idem k liječniku.
- 3) Idem kod liječnika.

Neizvorni govornici nemaju izvornogovorničku sposobnost koja bi im pomogla da spontano znaju kojim se riječima dopunjuje koji glagol u kojem značenju te im je zbog toga potrebno osvijestiti što je to valentnost te pružiti i strukturirati podatke o njoj, kao i proizvesti didaktičke uvjete u kojima će ih oni biti u mogućnosti usvojiti (Gulešić-Machata, Čilaš-Mikulić, Udier 2011: str. 24.).

Iz navedenih primjera možemo zaključiti koliko ovakva analiza valentnosti, a posebice izrada kompletnog valencijskog rječnika glagola u danom jeziku, može pridonijeti svakom korisniku jezika, kako izvornom govorniku, tako i neizvornom govorniku koji tek uči jezik.

Žabokrtský, jedan od autora VALLEX-a, u svojoj doktorskoj disertaciji ističe i važnost sintaktičkog leksikona glagola s informacijama o valentnosti za mnoge zadatke u domeni NLP-a<sup>1</sup>, odnosno obrade prirodnog jezika:

- *Lematizacija* (biranje ispravne leme za svaku riječ u tekstu)

1. Stali se matematiky.

Postali su matematičari.

---

<sup>1</sup> Natural Language Processing

## 2. Báli se matematiky.

Bojali su se matematike.

U obje rečenice se pojavljuje oblik *matematiky*, koji može biti akuzativ množine ili instrumental množine leme *matematik* (hr. matematičar) ili pak genitiv jednine, nominativ množine ili akuzativ množine leme *matematika* (hr. matematika). U obje se rečenice lema može razaznati pomoću činjenice da glagol *stát se* iz prve rečenice u svom valencijskom okviru ne sadrži akuzativ, a glagol *bát se* iz druge rečenice u svom valencijskom okviru također ne sadrži akuzativ.

- *Označivanje* (biranje ispravne morfološke oznake za određenu riječ i lemu)

## 3. Ptala se jeho bratra.

Pitala je svog brata.

Imenička sintagma *jeho bratra* može biti u genitivu jednine ili u akuzativu jednine, no glagol *ptát se* (hr. pitati) dopušta samo prvu mogućnost.

- *Sintaktička analiza* (određivanje koja riječ ovisi o kojoj riječi)

## 4. Nechala ho spát.

Pustila ga je da spava.

## 5. Začala ho milovat.

Počela ga je voljeti.



U četvrtoj rečenici zamjenica *ho* može ovisiti jedino o glagolu koji joj prethodi (budući da valencijski okvir glagola *spát* (hr. spavati) ne sadrži akuzativ), dok u petoj rečenici ista zamjenica mora ovisiti o glagolu koji slijedi iza nje (budući da za glagol *začít* (hr. početi) ne postoji okvir koji sadrži i akuzativ i infinitiv). Ukoliko bi se u obzir uzele samo morfološke oznake, obje rečenice bile bi ekvivalentne. Nedvosmislena ovisnosna struktura ne može se izraditi ako se ne uzmu u obzir valencijski okviri danih glagola.

- *Razrješavanje višeznačnosti riječi*

6. Odpovídal na otázky.

Odgovarao je na pitanja.

7. Odpovídal za děti.

Odgovarao je (bio je odgovoran) za djecu.

8. Odpovídal popisu.

Odgovarao je (bio je sukladan) opisu.

Različita značenja iste riječi često odgovaraju promjeni u valencijskim okvirima.

- *Semantička analiza*

9. Přišel po Petrovi.

Došao je nakon Petra.

10. Sháněl se po Petrovi.

Tražio je Petra.

Prijedložni izrazi najčešće imaju ulogu priložne oznake (kao u 9. rečenici), međutim mogu predstavljati i glagolski dodatak, što čini značajnu razliku u većini semantički ili logički motiviranih pristupa. Uloga prijedložnog izraza ne može se odrediti ako se ne uzmu u obzir valencijski okviri danih glagola.

- *Strojno prevođenje*

Svaki od gore navedenih problema neizbježno se pojavljuju prilikom bilo kakvog ozbiljnog pokušaja strojnog prevođenja (MT<sup>2</sup>). Postojanje valencijskog rječnika poboljšalo bi kvalitetu odgovarajućih podmodula takvog MT sustava, što bi u konačnici poboljšalo i kvalitetu samog prijevoda. (Žabokrtský 2005: str. 5.).

---

<sup>2</sup> Machine Translation

## V. Valencijski rječnici češkog jezika

Postoji više valencijskih rječnika češkog jezika, ali u ovom radu će biti spomenuti samo najvažniji za ovu temu, i to VALLEX u sve tri objavljene verzije.

### V.1. VALLEX 1.0

VALLEX (Valency Lexicon of Czech Verbs) je valencijski rječnik glagola u češkom jeziku. Kao što je spomenuto ranije, utemeljen je na projektu FGD (Functional Generative Description), koji se počeo razvijati u 60-im godinama 20. stoljeća i trenutno postoje tri verzije rječnika: Vallex 1.0, Vallex 2.0 i Vallex 2.5. Prva verzija ovog rječnika, Vallex 1.0, koji sadrži oko 1400 glagola, u ovom je radu najdetaljnije obrađena, budući da u najvećoj mjeri odgovara hrvatskoj inačici, CROVALLEX-u. Struktura podataka detaljno je obrađena u sljedećem poglavlju.

Vallex 2.0 i Vallex 2.5 su novija izdanja, koja su proširena i u teoretskom i kvantitativnom smislu.

Cilj ovog leksikona je postati elektroničkim leksikonom koji bi obuhvatio ne samo strukturu svih čeških glagola, nego i sintaktičko-semantičke informacije koje su nužne u računalnoj analizi i sintezi tekstova na češkom jeziku i svim drugim poznatim zadacima obrade prirodnog jezika. Valencijski leksikon čeških glagola je pokušaj prikupljanja i dokumentiranja skupa lingvistički označenih podataka čega je rezultat formalni opis valencijskih okvira glagola u češkom jeziku. Strukturu leksikona čine valencijskih okviri najučestalijih glagola u češkom jeziku, koji su unutar okvira podijeljeni s obzirom na značenje glagola. Okviri nude broj dopuna i njihove morfološke značajke te oznaku obveznosti odnosno izbornosti. Osim toga, valencijski okvir sadrži i primjere upotrebe pojedinog značenja glagola te dodatna obilježja kao što su idiomi, kontrola, refleksivnost, recipročnost i sintaktičko-semantička klasa.

## V.2. VALLEX 1.0. – logička struktura podataka

Terminologija korištena pri označavanju potječe većinom iz FGD-a ili se radi o općeprihvaćenim lingvističkim pojmovima.

Budući da je rječnik prilagođen korisnicima koji nisu izvorni govornici češkog jezika, sve oznake pojedinih strukturalnih jedinica nose nazive na engleskom jeziku. U svrhu lakšeg povezivanja ovdje opisanog gradiva sa samim rječnikom, u ovom radu će se također koristiti termini na engleskom jeziku uz njihove pripadajuće prijevode na hrvatski jezik.

### 1. Word entries (natuknice)

Na najopćenitijoj razini VALLEX 1.0 je podijeljen na natuknice. Svaka natuknica odgovara jednoj ili više lema. Svaka se natuknica sastoji od niza okvira za pojedinu lemu, a svaki okvir uglavnom odgovara jednom od značenja leme. Informacija o vidu lema je dodijeljena svakoj natuknici kao cjelini.

U većini slučajeva natuknice odgovaraju lemapa u smislu jedan-na-jedan, međutim VALLEX 1.0 bilježi i složenije oblike varijanata lema i homografa, o kojima će više riječi biti u točkama 3. i 4.

### 2. Lemmas (Leme)

Pod pojmom leme podrazumijevamo infinitivni oblik određenog glagola. U slučaju postojanja homografa, lemi je dodijeljen broj u superscriptu, a taj se homograf u VALLEX-u 1.0 smatra nerazdvojjim dijelom leme.

Povratne zamjenice *se* ili *si* smatraju se dijelom infinitiva samo kada je glagol pravi *reflexiva tantum* (npr. *bát se*, *smát se*) ili povratni glagol koji je izveden od nepovratnog oblika, ali je njegovo značenje toliko različito da se ne može smatrati povratnim oblikom originalnog leksema i da zapravo tvori zaseban leksem (npr. *hodit* / *hodit se*, *chovat* / *chovat se*).

### 3. Lemma Variants (varijante lema)

Varijante lema su grupe od dviju ili više lema, koje se u bilo kojem kontekstu mogu izmjenjivati bez promijene značenja (npr. *dovédět se/dozvédět se*). U većini slučajeva se radi o manjoj morfološkoj alternaciji, koja ponekad može donositi suptilan stilistički pomak (npr. *myslet/myslit*, pri kojem je drugi oblik „knjiški“). Premda se infinitivni oblici navedenih varijanti pravopisno razlikuju, njihovi su sklonidbeni oblici često identični (imp. sg. *mysli* vrijedi za oba oblika *myslet i myslit*).

Termin varijante lema ne bi trebao se miješati sa sinonimima.

### 4. Homonyms (homografi)

VALLEX 1.0 bilježi i parove natuknica čije leme imaju identičan pravopisni oblik, ali njihova se značenja bitno razlikuju. Mogu se također razlikovati u etimologiji (npr. *nakupovat*<sup>II</sup> – kupovati u odnosu na *nakupovat*<sup>I</sup> – gomilati), vidu (npr. *stačit*<sup>II</sup> pf. – biti dovoljan u odnosu na *stačit*<sup>I</sup> impf. – stići) te sklonidbenim oblicima (*žilo* (past.sg.fem) od *žít*<sup>I</sup> – živjeti u odnosu na *žalo* (past.sg.fem.) od *žít*<sup>I</sup> – kositi, žeti). Takve se leme (homografi) razlikuju pomoću numeriranja rimskom brojkom u superscriptu.

### 5. Frame Entries (Okviri) (leksičke jedinice)

Svaka se natuknica sastoji od niza okvira, koji uglavnom odgovaraju pojedinim značenjima (smislovima) lema. S ovog se gledišta VALLEX 1.0 može klasificirati kao *Sense Enumerated Lexicon* (leksikon numeriran prema značenjima).

U VALLEX-u 1.0 su okviri nabrojani unutar svake natuknice, a brojevi okvira su pridodani lemapa kao *subscript*.

Redoslijed lema nije sasvim nasumičan, ali nije niti savršeno sistematičan. Za sada je redoslijed određen na temelju čestote pojedinih pojava tako da se češća značenja nalaze na prvim mjestima, dok se rijetka ili idiomatska značenja nalaze na začelju liste. Autori naglašavaju da se njihov poredak temelji na njihovoj intuiciji te ne garantiraju da redoslijed odgovara čestoti njihovog pojavljivanja u suvremenom jeziku.

Svaki okvir sadrži opis samog valencijskog okvira te njegovih dodataka.

#### 6. Valency Frames (valencijski okviri)

U VALLEX-u 1.0 je valencijski okvir oblikovan kao niz polja. Svako polje okvira odgovara jednom komplementu (dopuni ili dodatku) danog glagola.

Svakom polju su pridijeljena sljedeća obilježja: funktor (7.), popis mogućih morfoloških oblika (8) te vrstu komplementa (11).

Autori su također primijetili da neka polja imaju tendenciju da se pojavljuju zajedno te su, kako bi zabilježili ovaj tip pravilnosti, razvili mehanizam proširivanja polja (12). U tim se slučajevima puni valencijski okvir dobiva nakon primijene tih proširenja

#### 7. Functors (funktori)

U VALLEX-u 1.0 se funktori (oznake za dubinske padeže) koriste za izražavanje tipova odnosa među glagolima i elementima koji dolaze uz njih. Prema FGD-u, funktori se dijele na *dopune* (inner participants) i *dodatke* (free modifications). VALLEX 1.0 također definira i dodatnu grupu kvazi-valentnih elemenata.

Funktori, koji se pojavljuju u VALLEX-u 1.0:

##### a) Inner participants (dopune):

ACT (actor) - Peter read a letter.

ADDR (addressee) - Peter gave Mary a book.

PAT (patient) - I saw him.

EFF (effect) - We made her the secretary.

ORIG (origin) - She made a cake from apples.

## b) Quasi-valency complementations (kvazi-valentni komplementi):

DIFF (difference) - The number has swollen by 200.

OBST(obstacle) - The boy stumbled over a stumb.

INTT (intent) - He came there to look for Jane.

## c) Free modifications (dodaci):

ACMP (accompagnement) - Mother came with her children.

AIM (aim) - John came to a bakery for a piece of bread.

BEN (benefactive) - She made this for her children.

CAUS (cause) - She did so since they wanted it.

COMPL (complement) - They painted the wall blue.

DIR1 (direction-from) - He went from the forest to the village.

DIR2 (direction-through) - He went through the forest to the village.

DIR3 (direction-to) - He went from the forest to the village.

DPHR (dependent part of a phraseme) - Peter talked horse again.

EXT (extent) - The temperatures reached an all time high.

HER (heritage) - He named the new villa after his wife.

LOC (locative) - He was born in Italy.

MANN (manner) - They did it quickly.

MEANS (means) - He wrote it by hand.

NORM (norm) - Peter has to do it exactly according to directions.

RCMP (recompense) - She bought a new shirt for 25 \$.

REG (regard) - With regard to George she asked his teacher for advice.

RESL (result) - Mother protects her children from any danger.

SUBS (substitution) - He went to the theatre instead of his ill sister.

TFHL (temporal-for-how-long) - They interrupted their studies for a year.

TFRWH (temporal-from-when) - His bad reminiscences came from this period.

THL (temporal-how-long ) - We were there for three weeks.

TOWH (temporal-to when) - He put it over to next Tuesday.

TSIN (temporal-since-when) - I have not heard about him since that time.

TWHEN (temporal-when) - His son was born last year.

Osim gore navedenih funktora, u VALLEX-u 1.0 se pojavljuje i oznaka DIR, koja se koristi kao poseban simbol za proširenje polja. Neki od funktora definiranih u projektu FGD su ovdje izostavljeni jer se ili uopće ne pojavljuju uz glagole ili ne utječu na valencijske okvire.

## 8. Morphemic forms (morfemski oblici)

U rečenici se svako polje okvira može izraziti određenim skupom morfemskih sredstava, koje nazivamo morfemskim oblicima. U VALLEX-u 1.0 se skup mogućih oblika definira ili eksplicitno ili implicitno. U prvom su slučaju oblici pobrojani u listi danog polja, dok u drugom slučaju takva lista nije određena jer je skup mogućih oblika određen funktorom danog polja, tj. mogu se pojaviti svi oblici koji mogu izražavati dani funktor.



## 9. Explicitly declared forms (eksplicitno određeni oblici)

Lista oblika dodijeljenih polju okvira može sadržavati vrijednosti sljedećih tipova:

- Čisti padeži (bez prijedloga) – u češkom jeziku, kao i u hrvatskom, postoji sedam morfoloških padeža. U VALLEX-u 1.0 autori koriste njihove tradicionalne brojčane oznake: 1 – nominativ, 2 – genitiv, 3 – dativ, 4 – akuzativ, 5 – vokativ, 6 – lokativ te 7 – instrumental.
- Padeži s prijedlozima – u ovim je slučajevima navedena lema prijedloga te broj padeža koji se koristi s danim prijedlogom (npr. z+2, na+4, o+6...). Prijedlozi koji se pojavljuju u VALLEX-u 1.0 su: bez, do, jako, k, kolem, kvůli, mezi, místo, na, nad, na úkor, o, od, ohledně, okolo, oproti, po, pod, podle, pro, proti, před, přes, při, s, u, v, ve prospěch, vůči, v zájmu, z, za. Iako se „jako“ uglavnom smatra veznikom, uključen je u ovu listu budući da u nekim valencijskim okvirima treba određeni padež.
- Veznici zavisno složenih rečenica – navodi se lema veznika. U VALLEX-u 1.0 se pojavljuju sljedeći tipovi veznika zavisno složenih rečenica: aby, ať, až, jak, zda, že.
- Infinitivne konstrukcije – kratica „inf“ – predstavlja element u obliku glagolskog infinitiva. Oznaka „inf“ se može pojaviti i zajedno sa prijedlogom (npr. „než+inf“), ali takvi su slučajevi u češkom jeziku rijetki.
- Konstrukcije s priložima – kratica „adj-broj“ se navodi za elemente u obliku priloga u određenom padežu, npr. adj-1 (cítím se slabý – osjećám se slabim).
- Konstrukcije s glagolom „být“ – infinitiv glagola „být“ (biti) može se naći u kombinaciji s nekim od gore navedenih tipova, npr. být+adj-1 (npr. zdá se to být dostatečně – izgleda da je to dovoljno).

- Dio frazema – ako je skup mogućih leksičkih vrijednosti danog komplementa vrlo malen (često samo jedan element), autori tada izravno nabroje te vrijednosti (npr. „napospas“ za frazem „ponechat napospas“ – ostaviti na milost i nemilost).

#### 10. Implicitly declared forms (implicitno određeni oblici)

Ukoliko za polje okvira oblici nisu eksplicitno navedeni, tada lista mogućih oblika implicitno proizlazi iz funktora polja i to prema sljedećem popisu:

- LOC: prilog, na+6, v+6, u+2, před+7, za+7, nad+7, pod+7, okolo+2, kolem+2, při+6, vedle+2, mezi+7, mimo+4, naproti+3, podél+2 ...
- MANN: prilog, 7, na+4, ...
- DIR3: prilog, na+4, v+4, do+2, před+4, za+4, nad+4, pod+4, vedle+2, mezi+4, po+4, okolo+2, kolem+2, k+3, mimo+4, naproti+3 ...
- DIR1: prilog, z+2, od+2, zpod+2, zpoza+2, zpřed+2 ...
- DIR2: prilog, 7, přes+4, podél+2, mezi+7, ...
- TWHEN: prilog, 2, 4, 7, před+7, za+4, po+6, při+6, za+2, o+6, k+3, mezi+7, v+4, na+4, na+6, kolem+2, okolo+2, ...
- THL: prolog, 4, 7, po+4, za+4, ...
- EXT: prilog, 4, na+4, kolem+2, okolo+2, ...
- REG: prilog, 7, na+6, v+6, k+3, při+6, ohledně+2, nad+7, na+4, s+7, u+2, ...
- TFRWH: z+2, od+2, ...
- AIM: k+3, na+4, do+2, pro+4, proti+3, aby, at', že, ...
- TOWH: na+4 ...
- TSIN: od+2 ...
- TFHL: na+4, pro+4 ...

- NORM: podle+2, v duchu+2, po+6, ...
- MEANS: 7, v+6, na+6, po+6, z+2, že, s+7, na+4, za+4, pod+7, do+2, ...
- CAUS: 7, za+4, z+2, kvůli+2, pro+4, k+3, na+4, že, ...

### 11. Types of complementations (vrste komplemenata)

U okviru projekta FGD, valencijski okviri u užem smislu riječi sastoje se samo od dopuna (obveznih i izbornih, skraćeno označenih s „*obl*“ i „*opt*“) te obveznih dodataka. U VALLEX-u 1.0 valencijskim su okvirima dodani i kvazi-valentni komplementi. Nadalje, u valencijskim se okvirima pojavljuju i neki neobvezni dodaci, budući da su tipično („*typ*“) povezani s nekim glagolima (ili cijelim klasama glagola), dok s drugima nisu.

Atribut „*type*“ je dodijeljen svakom polju okvira te može imati vrijednosti „*obl*“ ili „*opt*“ za glagolske dopune i kvazi-valentne komplemente, i oznake „*obl*“ ili „*typ*“ za glagolske dodatke.

### 12. Slot expansion (proširenje polja)

Neka polja imaju tendenciju da se sistematično pojavljuju zajedno. Primjerice, glagoli kretanja mogu često biti određeni smjerom (DIR1 – početna točka kretanja, DIR2 – među-točka kretanja ili DIR3 – ciljna točka kretanja). Ovaj tip pravilnosti autori su odlučili zabilježiti i označiti zasebnom oznakom za polje. Ukoliko postoji takva oznaka (u VALLEX-u 1.0 se za ovu oznaku koristi uspravna strelica), potpuni će se valencijski okvir dobiti nakon proširenja polja.

Dakle, ukoliko je jedno polje okvira obilježeno uspravnom strelicom, tada će se potpuniji valencijski okvir dobiti nakon što takvo polje zamijenimo prema sljedećim pravilima:

$$\uparrow \text{DIR}^{\text{typ}} \rightarrow \text{DIR1}^{\text{typ}} \text{DIR2}^{\text{typ}} \text{DIR3}^{\text{typ}}$$

$$\uparrow \text{DIR1}^{\text{obl}} \rightarrow \text{DIR1}^{\text{obl}} \text{DIR2}^{\text{typ}} \text{DIR3}^{\text{typ}}$$

$$\uparrow \text{DIR2}^{\text{obl}} \rightarrow \text{DIR1}^{\text{typ}} \text{DIR2}^{\text{obl}} \text{DIR3}^{\text{typ}}$$

$\uparrow\text{DIR3}^{\text{obl}}$   $\rightarrow$   $\text{DIR1}^{\text{typ}}$   $\text{DIR2}^{\text{typ}}$   $\text{DIR3}^{\text{obl}}$

$\uparrow\text{TSIN}^{\text{obl}}$   $\rightarrow$   $\text{TSIN}^{\text{obl}}$   $\text{THL}^{\text{typ}}$   $\text{TTIL}^{\text{typ}}$

$\uparrow\text{THL}^{\text{typ}}$   $\rightarrow$   $\text{TSIN}^{\text{typ}}$   $\text{THL}^{\text{typ}}$   $\text{TTIL}^{\text{typ}}$

### 13. Frame attributes (Atributi okvira)

Atributi okvira mogu u VALLEX-u 1.0 biti obvezni ili izborni. Obvezni atributi moraju biti popunjeni u svakom okviru, dok su izborni prazni u slučajevima kada ne postoje (npr. neki glagoli nemaju vidne parnjake) ili zbog nedovršenog označivanja (npr. atribut *class* je popunjen otprilike kod samo trećine okvira).

Obvezni atributi okvira su:

- *gloss* – glagol ili parafraza koja je otprilike sinonim danog okvira ili značenja. Ovaj atribut ne sadrži nužno stvarni sinonim, već pomaže u opisivanju značenja i u brzem snalaženju među natuknicama.
- *example* – potpuna rečenica ili dijelovi rečenica koje sadrže dani glagol u danom valencijskom okviru.

Izborni atributi okvira:

- *control* (opisan u 14. točki)
- *class* (opisan u 15. točki)
- *vidni parnjaci* (opisan u 16. točki)
- *idiom* (opisan u 17. točki)

### 14. Control (kontrola)

Termin *control* u ovom se kontekstu odnosi na određene tipove predikata (glagole kontrole) i dvije moguće koreferentne uloge: vršitelj kontrole i trpitelj kontrole.

## 15. Class (klasa)

Nekim su okvirima dodijeljene semantičke klase glagola tipa „kretanje“, „razmjena“, „komunikacija“, „percepcija“ itd. Doduše, kod ove oznake autori napominju da je klasifikacija probna i da bi se trebala shvatiti tek kao intuitivno grupiranje okvira. Ovaj tip klasifikacije je uveden u VALLEX-u 1.0 zbog činjenice da pojednostavljuje semantičku provjeru dosljednosti i omogućava promatranje podataka na općenitijoj razini.

## 16. Aspect, Aspectual counterparts (vid i vidni parnjaci)

U češkom jeziku, kao i hrvatskom, postoji razlika između svršenih glagola (u VALLEX-u 1.0 označenih kraticom *pf.* – perfective) i nesvršenih glagola (označenih kraticom *impf.* – imperfective). U VALLEX-u 1.0 ova se karakteristika navodi za cijelu natuknicu, tj. vrijednost je ista za sve njezine okvire i dijele je sve postojeće varijante lema.

Neki se glagoli (npr. *informovat*, *charakterizovat*) u različitim kontekstima mogu koristiti i kao svršeni i kao nesvršeni glagoli. Takvi glagoli se nazivaju dvovidnim glagolima i u VALLEX-u 1.0 su označeni kraticom *biasp.*

Među nesvršenim glagolima razlikujemo i iterativne (*iter.*) glagole, koji se u češkom uglavnom dobivaju umetanjem sufiksa *-va-* ili *-iva-*, a izražavaju radnje ponavljanja ili radnje produženog trajanja (npr. *čítávat*, *chodívat*). Iterativni glagoli koji sadrže dvostruki sufiks *-va-* (npr. *chodivávat*) su potpuno zanemareni, dok se ostali iterativni glagoli pojavljuju kao vidni parnjaci u okvirima odgovarajućih neiterativnih glagola.

Glagol u određenom značenju može imati svoj vidni parnjak koji ima gotovo jednako značenje s iznimkom svojstva vida. Na tektogramatičkoj razini FGD-a oba člana vidnog para tvoje jednu leksičku jedinicu, no u VALLEX-u 1.0 svaki od njih čini zasebnu natuknicu jer imaju različite morfološke oblike. Vidni parnjaci ne moraju nužno biti jednaki za sva značenja danog glagola, što je vidljivo kod primjera *odpovědět*, koji je vidni parnjak glagola *odpovídat* u značenju „odgovarati na pitanje“, međutim nije vidni parnjak istog glagola u značenju „odgovarati, pristajati nečemu“.

Jedan okvir može imati više od jednog vidnog parnjaka ako za njega postoje dva parnjaka istog vida (impf. *působit* i impf. *způsobovat* za pf. *způsobit*) ili ako postoje dva parnjaka različitog vida (impf. *scházet*, pf. *sejít*, iter. *schazívát*).

#### 17. Idiomatic frames (Idiomatski okviri)

Prilikom razvoja VALLEX-a 1.0 autori su se uglavnom usredotočili na glavna ili uobičajena značenja glagola. Zabilježeni su i mnogi okviri koji odgovaraju manje uobičajenoj uporabi, no ti slučajevi ipak nisu iscrpno obrađeni. Ovaj su tip okvira autori nazvali idiomatskim i označili ga oznakom *idiom*. Idiomatski okvir karakterizira značajan pomak u značenju (u odnosu na glavno značenje) ili mali i striktno određen skup mogućih leksičkih vrijednosti u jednom od njegovih komplementa ili pak postojanje drugih tipova nepravilnosti ili anomalija.

**V.3. VALLEX 2.0 i VALLEX 2.5**

VALLEX 2.0 i VALLEX 2.5 sastoje se od 2730 natuknica i sadrže oko 6460 značenja glagola. Za razliku od prethodne verzije, VALLEX 2.0 počinje tretirati vidni parnjak kao jedan leksem. Kada bi se svršeni i nesvršeni glagoli brojali zasebno, opseg VALLEX-a 2.0 bi narastao na 4250 jedinica.

Kao što je ranije napomenuto, novije verzije VALLEX-a su, osim povećanim brojem glagola, proširene i u teoretskom smislu te primjerice razlikuju nekoliko dodatnih implicitno određenih morfoloških oblika, ili su pak prošireni oni koji su i prethodno definirani.

Prema popratnoj literaturi koja je na raspolaganju uz VALLEX 2.5 i koja opisuje strukturu samog rječnika ([http://ufal.mff.cuni.cz/vallex/2.5/doc/structure\\_en.html](http://ufal.mff.cuni.cz/vallex/2.5/doc/structure_en.html)), vidljivo je da je popis znatno proširen u odnosu na prvotnu verziju:

ACMP: bez+2, s+7, společně s+7, spolu s+7, v čele s+7, v souvislosti s+7, ve spojení s+7, včetně+2, ... ;

AIM: aby, at', do+2, k+3, na+4, o+4, pro+4, pro případ+2, proti+3, v zájmu+2, za+4, za+7, že, ... ;

BEN: 3, na+4, na účet+2, na úkor+2, na vrub+2, pro+4, proti+3, v+4, ve prospěch+2, v rozporu s+7, s+7, v zájmu+2, ... ;

CAUS: 7, aby, přílog, díky+3, jelikož, ježto, k+7, kvůli+3, na+4, na+6, na základě+2, nad+7, následkem+2, od+2, pod+7, pod nápor+2, pod tíhou+2, pod váhou+2, poněvadž, pro+4, proto, protože, v+6, v důsledku+2, v souvislosti s+7, vinou+2, vlivem+2, vzhledem k+3, z+2, z důvodu+2, za+4, za+7, zásluhou+2, že, ... ;

CRIT: 2, 7, dle+2, podle+2, na+6, na základě+2, po vzoru+2, přiměřeně+3, v+6, v duchu+2, v rozporu s+7, v souladu s+7, v souhlase s+7, v závislosti na+6, ve shodě s+7, ve smyslu+2, ve světle+2, z titulu+2, ... ;

DIR1: přílog, od+2, s+2, z+2, ze strany+2, zpod+2, zpoza+2, zpřed+2, ... ;

DIR2: 7, přílog, kolem+2, cestou+2, mezi+7, napříč+7, po+6, podél+2, přes+4, skrz+4, v+6, ...

DIR3: 7, prilog, do+2, do čela+2, k+3, kolem+2, mezi+4, mimo+4, na+4, na+6, nad+4, naproti+3, okolo+2, po+4, po+6, pod+4, proti+3, před+4, přes+4, směrem do+2, směrem k+3, směrem na+4, v+4, vedle+2, za+4, za+7, ... ;

EXT: 2, 4, 7, prilog, do+2, kolem+2, k+3, na+4, na+6, nad+4, okolo+2, po+6, pod+7, přes+4, v+4, z+2, za+4, ... ;

LOC: prilog, blízko+2, blízko+3, daleko+2, do+2, kolem+2, mezi+7, mimo+4, na+4, na+6, na úroveň+2, nad+7, naproti+3, nedaleko+2, okolo+2, po+6, po boku+2, poblíž+2, pod+7, podél+2, proti+3, před+7, přes+4, při+6, stranou+2, u+2, uprostřed+2, uvnitř+2, v+6, v čele+2, v oblasti+2, v rámci+2, v řadě+2, vedle+2, za+4, za+7, ... ;

MANN: 7, prilog, do+2, formou+2, na+4, na+6, nad+4, o+4, po+6, pod+7, proti+3, před+7, při+6, přes+4, s+7, v+4, v+6, v podobě+2, ve formě+2, vedle+2, z+2, za+4, za+7, jak, že, ... ;

MEANS: 7, prilog, cestou+2, díky+3, do+2, na+4, na+6, o+6, po+6, pod+7, pomocí+2, prostřednictvím+2, přes+4, s+7, s pomocí+2, v+6, z+2, za+4, skrz+2, za pomoci+2, že, , ... ;

REG: 7, prilog, bez ohledu na+4, bez zřetele k+3, k+3, kolem+2, na+4, na+6, na téma+2, nad+7, nezávisle na+6, o+6, ohledně+2, po+6, pro+4, před+7, při+6, s+7, se zřetelem k+3, se zřetelem na+4, s ohledem na+4, u+2, v+6, v otázce+2, v případě+2, v rámci+2, v souvislosti s+7, ve věci+2, ve vztahu k+3, vůči+3, vzhledem k+3, z+2, z hlediska+2, za+4, ... ;

SUBS: jménem+2, namísto+2, místo+2, výměnou za+4, za+4, ... ;

TFHL: prilog, do+2, na+4, po+2, pro+4, ... ;

TFRWH: z+2, od+2, ... ;

THL: 2, 4, 7, prilog, až, dokud, do+2, na+4, po+4, po dobu+2, přes+4, v+2, za+4, ... ;

TOWH: prilog, do+2, k+3, na+4, pro+4, ... ;

TSIN: prilog, od+2, počínaje+7, z+2, ... ;

TTILL: prilog, do+2, dokud, k+3, než, po+4, ... ;

TWHEN: 2, 4, 7, prilog, až, do+2, jakmile, k+3, když, kolem+2, koncem+2, mezi+7, na+4, na+6, na závěr+2, než, o+6, okolo+2, po+6, počátkem+2, postupem+2, poté co,



před+7, předtím než, při+6, s+7, u příležitosti+2, v+4, v+6, v době+2, v období+2, v průběhu+2, v závěru+2, z+2, za+2, za+4, začátkem+2, ... ;

Kao i u prvoj verziji VALLEX-a glagoli se razlikuju i oznakom povratnosti, koja je pak proširena s nekoliko novih oznaka za slučajeve kada je povratna zamjenica dio izvedenog pasivnog oblika glagola:

- za prijelazne glagole (npr. *plány se připravují* – planovi se pripremaju) uvedena je oznaka *pass*.
- za neprijelazne glagole (npr. *v neděli se chodí do kostela* – nedjeljom se ide u crkvu) uvedena je oznaka *pass0*.

Kasnije verzije VALLEX-a uvode i oznake za slučajeve kada povratne zamjenice *se* ili *si* zamjenjuju dopunu koja je u koreferentnom odnosu sa subjektom ACT. Oznaka *cor3* se koristi za označavanje koreferentnog odnosa između ACT i ADDR u dativu (npr. *darovat si něco* – darovati si nešto), dok oznaka *cor4* označava koreferentnost između ACT i PAT u akuzativu (npr. *mýt se* – prati se).

Uvedeno je i razlikovanje na temelju uzajamnosti (reciprociteta), koja se shvaća kao mogućnost glagolskih komplementa da budu u međusobnom odnosu koji je simetričan i u kojem se uloge izmjenjuju. U se češkom jeziku (kao i u hrvatskom), ako su ACT i neki komplementi međusobno recipročni, koristi povratni oblik glagola, a komplementi se tada izražavaju imeničkim skupom (*Petr a Marie se hádali* – Petar i Marija su se svađali (jedno s drugim)) ili imenicom u množini (*přátelé se návštěvují* – prijatelji se međusobno posjećuju) te moguće i dodatnim prilogom *spolu, navzájem* (zajedno, uzajamno).

Mogućnost recipročnog korištenja je navedena u oznaci reciprociteta (skraćeno *rcp*).

U odnosu na VALLEX 1.0 u novijim su verzijama autori uveli 7 novih semantičkih klasa i u konačnici je 45% leksičkih jedinica razdijeljeno i prema semantičkim klasama, koje obuhvaćaju leksičke jedinice sličnih sintaktičkih osobina i semantičkog značenja. Autori su tako uveli ukupno 22 semantičke klase glagola, koje obuhvaćaju 2903 leksičke jedinice.

## VI. CROVALLEX - valencijski rječnik hrvatskog jezika

CROVALLEX je hrvatski valencijski leksikon glagola, koji je u sklopu doktorske disertacije *Pristupi izradi strojnog tezaurusa za hrvatski jezik* izradila doc. dr. sc. Nives Mikelić Preradović pod vodstvom prof. dr. sc. Damira Borasa na Filozofskom Fakultetu Sveučilišta u Zagrebu.

CROVALLEX se, kao i češki VALLEX, temelji na Sgallovom Funkcijskom Generativnom Opisu (FGD) te sadrži 1739 najučestalijih glagola preuzetih iz Hrvatskog čestotnog rječnika<sup>3</sup>. Autorica u svojoj doktorskoj disertaciji naglašava da je Hrvatski čestotni rječnik već u doba izrade valencijskog rječnika bio star 8 godina te da je upitno koliko se popis najučestalijih glagola do tada promijenio. Međutim, glagoli popisani u CROVALLEX-u svakako predstavljaju kvalitetnu osnovu za daljnji razvoj i proširenje, kako većim brojem glagola, tako i proširenjem logičke strukture samog rječnika.

Budući da ovakav rječnik, koji je javno dostupan i strojno čitljiv, prethodno nije postojao, primarna svrha tog projekta je bila upravo izrada rječnika koji bi poslužio i ostalim istraživačima.

Trenutni opseg rječnika je 1739 glagola s njihovih 5118 valencijskih okvira, a autorica očekuje da bi rječnik u budućim izdanjima mogao sadržavati dvostruko više glagola, biti proširen novim klasama glagola te podijeljen na skupove sinonima umjesto na individualne leme glagola (Mikelić-Preradović 2010: str. 53-57.).

## VII. Usporedbena analiza valentnosti glagola u češkom i hrvatskom jeziku

U sklopu ovog istraživačkog rada provedena je usporedbena analiza valencijskih okvira čeških i hrvatskih glagola koji dijele isto značenje. Analiza je provedena izravnom usporedbom valencijskih okvira koje pojedini glagoli vežu na sebe, a kao baza podataka korišteni su valencijski rječnici CROVALLEX te VALLEX 2.5.

---

<sup>3</sup> Moguš, M., Bratanić, M., Tadić, M. Hrvatski čestotni rječnik. Zagreb: Zavod za lingvistiku i Školska knjiga. 1999.

Budući da oba jezika pripadaju skupini slavenskih jezika te zahvaljujući velikoj sličnosti u sintaksi, pronađen je velik broj glagolskih parnjaka čiji su valencijski okviri identični. Uočene su dakako i brojne razlike, od kojih su neke kategorizirane i popisane u nastavku ovog rada, dok su izostavljene razlike u valencijskim okvirima kod kojih poveznice ili uopće ne postoje ili nisu očite i dosljedne. Rezultati su u ovom radu podijeljeni u nekoliko skupina:

#### 1) Glagoli dijele isti valencijski okvir

1a) U najvećem su broju, dakako, primijećeni glagoli koji se mogu prevesti direktno, tj. kod kojih i češka i hrvatska inačica dijele isti valencijski okvir. Ovakva je situacija vrlo učestala kod glagola koji na sebe vežu vršitelja radnje (ACT) u nominativu te trpitelja radnje (PAT) u akuzativu, no nađeni su i drugi primjeri:

- analyzovat / analizirati: ACT\_1\_obl PAT\_4\_obl
- absorbovat / apsorbirati: ACT\_1\_obl PAT\_4\_obl
- bombardovat / bombardirati: ACT\_1\_obl PAT\_4\_obl
- napodobovat / oponašati: ACT\_1\_obl PAT\_4\_obl
- ranit / raniti: ACT\_1\_obl PAT\_4\_obl
- rodit / roditi: ACT\_1\_obl PAT\_4\_obl
- kouřit / pušiti: ACT\_1\_obl PAT\_4\_obl
- uklízet / čistiti: ACT\_1\_obl PAT\_4\_obl
- umět / znati: ACT\_1\_obl PAT\_4\_obl

#### 1b) ostali primjeri:

- bít (tlouci) / biti (udarati): ACT\_1\_obl PAT\_4\_obl + MEANS\_7\_typ
- bdít / bdjeti: ACT\_1\_obl PAT\_nad+7\_obl
- hodit / baciti: ACT\_1\_obl ADDR\_3\_opt PAT\_4\_obl
- porovnat / usporediti: ACT\_1\_obl PAT\_4\_obl EFF\_s+7\_obl
- klást / stavljati: ACT\_1\_obl PAT\_4\_obl DIR3\_obl

- oddělit / odijeliti: ACT\_1\_obl PAT\_4\_obl ORIG\_od+2\_obl

- ozdobit / okтити: ACT\_1\_obl PAT\_4\_obl INST\_7\_typ

itd.

2) Glagoli koji se razlikuju u valencijskom okviru, tj. vežu na sebe različit broj obaveznih dopuna.

2.1) Kao najčešći i najdosljedniji primjeri pokazali su se slučajevi u kojima glagoli na sebe veže vršitelja radnje (ACT) u nominativu, primatelja radnje (ADDR) u dativu te trpitelja radnje (PAT) u akuzativu. U ovim su slučajevima češkom glagolu sve tri dopune obavezne, dok je hrvatskom glagolu primatelj radnje tipična izborna dopuna:

- slíbit ACT\_1\_obl **ADDR\_3\_obl** PAT\_4,inf,zda,že,cont\_obl

- obećati ACT\_1\_obl **ADDR\_3\_typ** PAT\_inf,4\_obl

- vysvětlit ACT\_1\_obl **ADDR\_3\_obl** PAT\_4,aby,at',zda,že,cont\_obl

- objasniti ACT\_1\_obl **ADDR\_3\_typ** PAT\_4\_obl

- ulehčit ACT\_1\_obl **ADDR\_3\_obl** PAT\_4\_obl MEANS\_7\_typ

- olakšati ACT\_1\_obl **ADDR\_3\_typ** PAT\_4\_obl MEANS\_7\_typ

- darovat ACT\_1\_obl **ADDR\_3\_obl** PAT\_4\_obl CAUS\_k+3\_typ

- darovati ACT\_1\_obl **ADDR\_3\_typ** PAT\_4\_obl CAUS\_k+3\_typ

- radit ACT\_1\_obl **ADDR\_3\_obl** PAT\_4,inf,jak,zda,že,cont\_obl

- savjetovati ACT\_1\_obl **ADDR\_3\_typ** PAT\_4\_obl

- umožnit ACT\_1\_obl **ADDR\_3\_obl** PAT\_4,inf,aby,že\_obl

---

- omogućiti	ACT_1_obl <b>ADDR_3_typ</b> PAT_4_obl
- odpustit	ACT_1_obl <b>ADDR_3_obl</b> PAT_4,že,cont_obl
- oprostiti	ACT_1_obl <b>ADDR_3_typ</b> PAT_4_obl
- upozornit	ACT_1_obl <b>ADDR_4_obl</b> PAT_na+4,aby,at',zda,že,cont_obl
- upozoriti	ACT_1_obl <b>ADDR_4_typ</b> PAT_na+4,sub-da,sub-kako_obl
- přesvědčit	ACT_1_obl <b>ADDR_4_obl</b> PAT_o+6,inf,aby,že,at'_obl
- uvjeriti	ACT_1_obl <b>ADDR_4_typ</b> PAT_sub-da,u+4_obl
- vrátit	ACT_1_obl <b>ADDR_3_obl</b> PAT_4_obl
- vratiti	ACT_1_obl <b>ADDR_3_typ</b> PAT_4_obl
- zakázat	ACT_1_obl <b>ADDR_3_obl</b> PAT_4,inf,aby_obl
- zabraniti	ACT_1_obl <b>ADDR_3_typ</b> PAT_4_obl
- popřát	ACT_1_obl <b>ADDR_3_obl</b> PAT_2,4,aby,at'_obl
- zaželjati	ACT_1_obl <b>ADDR_3_typ</b> PAT_4_obl
- obětovat	ACT_1_obl <b>ADDR_3_obl</b> PAT_4_obl
- žrtvovati	ACT_1_obl <b>ADDR_3_typ</b> PAT_4_obl

---

2.2) Dosljednost postoji i kod glagola kojima se izražava promjena mjesta vršitelja ili trpitelja radnje. U sljedećim primjerima je funktor DIR češkom glagolu obavezna dopuna, dok je hrvatskom glagolu tipična izborna dopuna:

- |                          |                                       |
|--------------------------|---------------------------------------|
| - posadit se             | ACT_1_obl <b>DIR3_obl</b>             |
| - sjesti                 | ACT_1_obl <b>DIR3_typ</b>             |
|                          |                                       |
| - odnést                 | ACT_1_obl PAT_4_obl <b>DIR1_obl</b>   |
| - odnijeti               | ACT_1_obl PAT_4_obl                   |
|                          |                                       |
| - odvést                 | ACT_1_obl PAT_4_obl <b>DIR3,1_obl</b> |
| - odvesti                | ACT_1_obl PAT_4_obl <b>DIR3_typ</b>   |
|                          |                                       |
| - osvobodit <sup>4</sup> | ACT_1_obl PAT_4_obl <b>DIR1_obl</b>   |
| - osloboditi             | ACT_1_obl PAT_4_obl                   |
|                          |                                       |
| - vzdálit                | ACT_1_obl PAT_4_obl <b>DIR1_obl</b>   |
| - udaljiti               | ACT_1_obl PAT_4_obl <b>DIR1_typ</b>   |
|                          |                                       |
| - vracet se              | ACT_1_obl <b>DIR3_obl</b>             |
| - vraćati se             | ACT_1_obl <b>DIR3_typ</b>             |

2.3) Ostali primjeri:

---

<sup>4</sup> pustiti na slobodu

- hlasovat	ACT_1_obl PAT_pro+4_obl
- glasovati	ACT_1_obl PAT_za+4_typ
- ptát se	ACT_1_obl ADDR_2_obl PAT_na+4,zda,cont_obl
- pitati	ACT_1_obl PAT_4_obl
- obvinít	ACT_1_obl ADDR_4_obl PAT_z+2,že_obl
- optužiti	ACT_1_obl PAT_4_obl CAUS_sub-da,za_typ
- osvobodit <sup>5</sup>	ACT_1_obl PAT_4_obl ORIG_od+2,z+2_obl
- osloboditi	ACT_1_obl PAT_4_obl ORIG_od+2_typ
- ukrást	ACT_1_obl PAT_4_obl DIR_1_obl ili ADDR_3_obl
- ukrasti	ACT_1_obl PAT_4_obl
- žádat	ACT_1_obl PAT_4,inf,aby,at',zda,že_obl ORIG_na+6,od+2,po+6_obl
- zahtíjovati	ACT_1_obl PAT_4,sub-da_obl ORIG_od+2_typ
- zamyslet se	ACT_1_obl PAT_nad+7_obl
- zamisliti se	ACT_1_obl PAT_nad+7_typ

---

<sup>5</sup> riješiti obaveze

- stěžovat si      ACT\_1\_obl ADDR\_3\_obl PAT\_na+4\_že,cont\_obl

- žaliti se      ACT\_1\_obl CAUS\_na+4\_typ

Kao što je vidljivo iz svih navedenih primjera, češki glagoli gotovo uvijek imaju više obaveznih dopuna nego hrvatski. Glagoli koji u hrvatskom vežu više dopuna nego češki pokazali su se iznimno rijetkima.

- poděkovat      ACT\_1\_obl PAT\_3\_obl RCMP\_za+4\_typ

- zahvaliti      ACT\_1\_obl REC\_3\_obl PAT\_za+4,na+6\_obl

3) Treća skupina glagola koja je obuhvaćena ovim istraživanjem su glagoli koji imaju jednak broj dopuna, ali su dopune u različitim padežima.

- blahopřát      ACT\_1\_obl PAT\_3\_obl CAUS\_k+3\_typ

- čestitati      ACT\_1\_obl BEN\_3\_typ RESL\_4\_obl

- bránit      ACT\_1\_obl PAT\_4\_obl EFF\_proti+3,před+7\_opt

- brániti      ACT\_1\_obl PAT\_4\_obl RESL\_od+2\_typ

- starat se      ACT\_1\_obl PAT\_o+4\_obl

- brinuti se      ACT\_1\_obl PAT\_o+6\_obl

- toužit      ACT\_1\_obl PAT\_po+6\_obl

- čeznuti      ACT\_1\_obl PAT\_za+4\_obl

- všimnout si      ACT\_1\_obl PAT\_2,jak,zda,že,cont\_obl

- opaziti      ACT\_1\_obl PAT\_4,kako,da\_obl



- zbláznit se	ACT_1_obl PAT_do+2_obl
- poludjeti	ACT_1_obl PAT_za+7_obl
- docílit	ACT_1_obl PAT_2,4_obl
- postići	ACT_1_obl PAT_4_obl
- počítat	ACT_1_obl PAT_s+7,že_obl
- računati	ACT_1_obl PAT_na+4,sub-da_obl
- rozumět	ACT_1_obl PAT_3_zda,že,cont_obl
- rozumjeti	ACT_1_obl PAT_4_obl
- zdát se	ACT_3_obl PAT_1,že,jak_obl
- sanjati	ACT_1_obl PAT_4_obl
- pamatovat se	ACT_1_obl PAT_na+4,jak,zda,že,cont_obl
- sjećati se	ACT_1_obl PAT_2_obl
- použít	ACT_1_obl PAT_2,4_obl AIM_k+3,na+4,aby_typ
- upotrijebiti	ACT_1_obl PAT_4_obl AIM_sub-da,za+4_typ

## VIII. Pravila realizacije dopuna

Tijekom analize uočeno je i nekoliko pravila po kojima se određene dopune čeških glagola realiziraju u hrvatskom jeziku:

1) Češka dopuna PAT\_o+4 se u hrvatskom najčešće realizira kao PAT\_za+4:

- bát se **o syna** / bojati se **za sina**
- bojovat **o zlatou medaili** / boriti se **za zlatnu medalju**
- hrát **o postup** do finále / igrati **za ulazak** u finale
- prosit/žádat bratra **o pomoc** / moliti brata **za pomoć**
- pečovat/starat se **o děti** / brinuti se **za djecu**
- soupeřit/soutěžit **o titul mistra** / natjecati se **za titulu prvaka**
- zajímat se **o historii** / zanimati se **za povijest**

2) Slična se pojava zamjećuje i kod dopune DIFF\_o+4, koja se u hrvatskom realizira kao DIFF\_za+4

- padat/růst **o 10%** / padati/rasti **za 10%**
- klesat/stoupat **o 10%** / padati/rasti **za 10%**
- zkrátit/prodloužit sukni **o 5 cm** / skratiti/produljiti suknju **za 5cm**
- obohatit se **o miliardu eur** / obogatiti se **za milijardu eura**
- odložit platbu **o 60 dní** / odgoditi plaćanje **za 60 dana**
- ohřát vodu **o 10 stupňů** / zagrijati vodu **za 10 stupnjeva**
- posunout zkoušku **o týden** / pomaknuti ispit **za tjedan dana**
- prohloubit koryto **o 1 metr** / produbiti korito **za 1 metar**
- prohrát/vyhrát **o jednu branku** / izgubiti/pobijediti **za 1 gol**
- protáhnout přednášku **o 10 minut** / produžiti predavanje **za 10 minuta**
- přesáhnout průměrnou úroveň **o 5 metrů** / premašiti prosječnu razinu **za 5 metara**
- převýšit bratra **o 5 centimetrů** / prerasti brata **za 5 centimetara**
- přežít muže **o 5 let** / nadživjeti muža **za 5 godina**
- roztáhnout boty **o pár centimetrů** / rastegnuti cipele **za par centimetara**
- snížit/zvětšit ceny **o 10%** / sniziti/povećati cijene **za 10%**

3) Funktor CAUS\_7 se u hrvatskom često mora prevesti funktorima CAUS\_od+2 ili CAUS\_zbog+2

- trpět **epilepsií** -> bolovati **od epilepsije**
- chvět se/trást se **zimou** -> tresti se **od zime**
- křičet, řvát **bolesti** -> vrištati, urlati **od boli**
- povzdechnout **únavou** -> uzdahnuti **od umora**
- zhroutit se vlastní **vahou** -> raspasti se **od vlastite težine**
- omluvit nepřítomnost **nemocí** -> ispříčati izostanak **zbog bolesti**
- umírat/zacházet **hladem** -> umirati **od gladi**
- unavit se **práci** -> umoriti se **od rada**
- zbláznit se **nedočkavosti** -> poludjeti **od nestrpljivosti**
- znepokojit se **špatnou zprávou** -> uznemíriti se **zbog loše vijesti**

4) Funktor CAUS\_z+2 također se ponekad prevodi kao CAUS\_od+2

- zbláznit se **z lásky** / poludjeti **od ljubavi**
- zhroutit se **z přepracování** -> slomiti se **od previše posla**
- rozplakat se **z lítosti** -> rasplakati se **od žalosti**
- vyléčit pacienta **z nachlazení** -> izliječiti pacijenta **od prehlade**

## IX. Zaključak

U ovom su radu opisani začeci Tesnièreove teorije valentnosti, njezin razvoj te konačno njezina primjena u izradi današnjih valencijskih rječnika. Valencijski rječnici VALLEX i CROVALLEX su dva vrlo vrijedna projekta koji mogu biti korisni izvornim govornicima hrvatskog i češkog jezika, ali i strancima koji ove jezike tek uče. Osim što pomažu u razumijevanju same problematike valencija i povezivanja leksema, ovi će rječnici pomoći i u ostvarenju većeg projekta: strojne analize rečenica. Ukoliko bi se na temelju ovih dvaju rječnika izradio dvojezični valencijski rječnik, on bi bio vrijedan korak prema kvalitetnom strojnom prevođenju.

S nadom da bi upravo to mogao biti početak tog zahtjevnog projekta, prilikom izrade ovog rada je provedeno istraživanje pravilnosti po kojima češki i hrvatski glagoli realiziraju svoje dopune i valencijske okvire. Budući da oba jezika pripadaju skupini slavenskih jezika te da postoje brojne sličnosti u samoj sintaksi ovih dvaju jezika, neka pravila su se pokazala vrlo uočljivim i vrlo dosljednim.

Iako je put do izrade potpunih valencijskih rječnika obaju jezika još uvijek poprilično dugačak, posebice ako imamo u vidu i valencijske rječnike imenica i pridjeva, ovakvi ambiciozni projekti nas doista dovode sasvim blizu ostvarenju želje mnogih stručnjaka u strojnoj obradi prirodnog jezika – razvoju inteligentnog i pouzdanog alata za strojno prevođenje.

## X. Literatura

1. Gulešić-Machata, M.; Čilaš-Mikulić, M.; Udier, S. L. Glagolske valencije i inojezični hrvatski, LAHOR – 11 (2011); Članci i rasprave; str. 23–38, 2011
2. Loptaková, M.; Žabokrtský Z. Valency Dictionary of Czech Verbs: Complex Tectogrammatical Annotation, Center for Computational Linguistics Faculty of Mathematics and Physics, Charles University, 2002
3. Lopatková, M.; Žabokrtský, Z.; Benešová, V., Valency Lexicon of Czech Verbs, 2006
4. Lopatková, M.; Žabokrtský, Z.; Benešová, V. VALLEX 1.0 Valency Lexicon of Czech Verbs, <http://ufal.mff.cuni.cz/vallex/1.0/> 02.08.2012
5. Lopatková, M.; Žabokrtský, Z.; Benešová, V., Valency Lexicon of Czech Verbs VALLEX 2.0, 2006
6. Lopatková, M.; Žabokrtský, Z.; Kettnerová, V. VALLEX 2.5 Valency Lexicon of Czech Verbs, <http://ufal.mff.cuni.cz/vallex/2.5/doc/home.html>, 01.08.2012
7. Mikelić-Preradović, N. Pristupi izradi strojnog tezaurusa za hrvatski jezik, doktorska disertacija, Odsjek za informacijske znanosti, Filozofski fakultet, Sveučilište u Zagrebu, 2008.
8. Mikelić-Preradović, N. Semantic classification of verbs in CROVALLEX // Recent advances in computer engineering and applications, 2010.
9. Vlastelić, A. Teorija valentnosti ogranak je gramatike zavisnosti, [http://www.ffri.uniri.hr/cro/dogadjanja/2003\\_2004/valentnost.htm](http://www.ffri.uniri.hr/cro/dogadjanja/2003_2004/valentnost.htm), 07.08.2012
10. Žabokrtský, Z. Valency Lexicon of Czech Verbs, doctoral thesis, Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, 2005

## **XI. Sažetak na hrvatskom jeziku**

U ovom je radu opisana kratka povijest nastanka valencijskih rječnika te same teorije valentnosti na kojoj se oni temelje. Potom su navedeni i opisani najznačajniji valencijski rječnici glagola u češkom jeziku te valencijski rječnik glagola u hrvatskom jeziku, koji je nastao na istim temeljima. Kao cilj rada je postavljeno utvrđivanje pravila po kojima se određeni valencijski okviri u češkom jeziku mogu direktno prevoditi na hrvatski jezik. Provedena je usporedbena analiza valencijskih okvira čeških i hrvatskih glagola koji dijele isto značenje. Za analizu su korištene dvije baze podataka: valencijski rječnici CROVALLEX 1.0 te VALLEX 2.5. Budući da oba jezika pripadaju skupini slavenskih jezika te zahvaljujući velikoj sličnosti u sintaksi, pronađen je velik broj glagola čiji su valencijski okviri identični. Uočene su dakako i brojne razlike, koje su kategorizirane i popisane u završnom dijelu rada.

Ključne riječi: valencija, valencijski rječnik, teorija valencije, valencijski okvir, analiza valencijskih okvira

## **XII. Sažetak na češkom jeziku**

V úvodu této práce je popsána krátká historie valenčních slovníků a teorie valence, na které se tyto slovníky zakládají. Dále jsou uvedeny a popsány nejvýznamnější valenční slovníky sloves v českém jazyce a valenční slovník sloves v chorvatském jazyce, který vznikl na stejných základech. Cílem práce bylo určit pravidla, podle kterých se určité valenční rámce v češtině mohou přímo přeložit do chorvatštiny. Také byla provedena srovnávací analýza valenčních rámců českých a chorvatských sloves, které mají stejný význam. K analýze byly používány dvě databáze: valenční slovníky CROVALLEX 1.0 a VALLEX 2.5. Vzhledem k tomu, že oba jazyky patří do skupiny slovanských jazyků, a díky velké podobnosti v syntaxi, bylo nalezeno velké množství sloves, jejichž valenční rámce jsou identické. Samozřejmě, bylo nalezeno i množství odlišností, které jsou kategorizovány a popsány v závěrečné části této diplomové práce.

Klíčová slova: valence, valenční slovník, teorie valence, valenční rámec, analýza valenčních rámců